Codebook

Duncan Turnbull

Sunday, December 21, 2014

CodeBook

Raw data activity measurements have been transformed into summaries in a tidy data format. The summary is IDs into Names

Data file

The data file is: tidyoutput.txt

It contains a header row defining the variables - explanations below It is 180 rows by 68 columns. The first two columns are the grouping columns: ActivityName and Subject The next 66 columns are the variables summarised by averagig for each subject by activity type.

e.g. (truncated for clarity)

```
> head(tidyoutput[1:7],2)
Source: local data frame [2 x 7]
Groups: ActivityName

  ActivityName Subject tBodyAccmeanX tBodyAccmeanY tBodyAccmeanZ tBodyAccstdX tBodyAccstdY
1       LAYING       1     0.2215982   -0.04051395    -0.1132036   -0.9280565   -0.8368274
2       LAYING       2     0.2813734   -0.01815874    -0.1072456   -0.9740595   -0.9802774
```

The activity subject is doing in observation 1 is LAYING and the subject id is 1 The tBodyAccmeanX value of 0.2215982 is the average of all the observations for subject 1 while LAYING.

Explanations

ActivityName: These are the names of the activities the subjects were completing during the observation (hopefully self explanatory). The values are:

- WALKING
- WALKING_UPSTAIRS
- WALKING_DOWNSTAIRS
- SITTING
- STANDING
- LAYING

Subject: This is the numeric id of the subject from whom the measurements were taken. Value range from 1 to 30

Every other variable in the output is an **average** of the measurements for the specific subject and the activity they were doing.

The variable names are made up of representative info as follows:

- These base measurements observed (before summarisation) were normalised and bounded within [-1,1].
- The observation summary is either time or frequency. This is represented with the first letter: t or f
- The end of the name will be std or mean then X,Y,Z or blank.
  - The X Y or Z letter represents the direction the signal reported.
  - The mean or std represents whether this summarised observations were means or standard deviations.
- The remaining label between these values represent the type of sensor, body or gyroscope, acceleration e.g. tGravityAccstdZ will show the average of the standard deviation of a time variable of Gravity Acceleration in the Z direction for the respective subject carrying out the respective activity

The columns are broken out for clarity a little more in this table:

```
> codebook
                    codes timefreq measure direction             feature
1            ActivityName                                   ActivityName
2                 Subject                                        Subject
3           tBodyAccmeanX        t    mean         x            BodyAcc
4           tBodyAccmeanY        t    mean         y            BodyAcc
5           tBodyAccmeanZ        t    mean         z            BodyAcc
6            tBodyAccstdX        t     std         x            BodyAcc
7            tBodyAccstdY        t     std         y            BodyAcc
8            tBodyAccstdZ        t     std         z            BodyAcc
9         tGravityAccmeanX        t    mean         x         GravityAcc
10        tGravityAccmeanY        t    mean         y         GravityAcc
11        tGravityAccmeanZ        t    mean         z         GravityAcc
12         tGravityAccstdX        t     std         x         GravityAcc
13         tGravityAccstdY        t     std         y         GravityAcc
14         tGravityAccstdZ        t     std         z         GravityAcc
15       tBodyAccJerkmeanX        t    mean         x        BodyAccJerk
16       tBodyAccJerkmeanY        t    mean         y        BodyAccJerk
17       tBodyAccJerkmeanZ        t    mean         z        BodyAccJerk
18        tBodyAccJerkstdX        t     std         x        BodyAccJerk
19        tBodyAccJerkstdY        t     std         y        BodyAccJerk
20        tBodyAccJerkstdZ        t     std         z        BodyAccJerk
21          tBodyGyromeanX        t    mean         x            BodyGyro
22          tBodyGyromeanY        t    mean         y            BodyGyro
23          tBodyGyromeanZ        t    mean         z            BodyGyro
24           tBodyGyrostdX        t     std         x            BodyGyro
25           tBodyGyrostdY        t     std         y            BodyGyro
26           tBodyGyrostdZ        t     std         z            BodyGyro
27      tBodyGyroJerkmeanX        t    mean         x       BodyGyroJerk
28      tBodyGyroJerkmeanY        t    mean         y       BodyGyroJerk
29      tBodyGyroJerkmeanZ        t    mean         z       BodyGyroJerk
30       tBodyGyroJerkstdX        t     std         x       BodyGyroJerk
31       tBodyGyroJerkstdY        t     std         y       BodyGyroJerk
32       tBodyGyroJerkstdZ        t     std         z       BodyGyroJerk
33         tBodyAccMagmean        t    mean                   BodyAccMag
34          tBodyAccMagstd        t     std                   BodyAccMag
35      tGravityAccMagmean        t    mean                GravityAccMag
36       tGravityAccMagstd        t     std                GravityAccMag
37     tBodyAccJerkMagmean        t    mean               BodyAccJerkMag
38      tBodyAccJerkMagstd        t     std               BodyAccJerkMag
39        tBodyGyroMagmean        t    mean                  BodyGyroMag
```

```
40          tBodyGyroMagstd        t     std                 BodyGyroMag
41     tBodyGyroJerkMagmean        t    mean             BodyGyroJerkMag
42      tBodyGyroJerkMagstd        t     std             BodyGyroJerkMag
43            fBodyAccmeanX        f    mean         x            BodyAcc
44            fBodyAccmeanY        f    mean         y            BodyAcc
45            fBodyAccmeanZ        f    mean         z            BodyAcc
46             fBodyAccstdX        f     std         x            BodyAcc
47             fBodyAccstdY        f     std         y            BodyAcc
48             fBodyAccstdZ        f     std         z            BodyAcc
49        fBodyAccJerkmeanX        f    mean         x        BodyAccJerk
50        fBodyAccJerkmeanY        f    mean         y        BodyAccJerk
51        fBodyAccJerkmeanZ        f    mean         z        BodyAccJerk
52         fBodyAccJerkstdX        f     std         x        BodyAccJerk
53         fBodyAccJerkstdY        f     std         y        BodyAccJerk
54         fBodyAccJerkstdZ        f     std         z        BodyAccJerk
55           fBodyGyromeanX        f    mean         x           BodyGyro
56           fBodyGyromeanY        f    mean         y           BodyGyro
57           fBodyGyromeanZ        f    mean         z           BodyGyro
58            fBodyGyrostdX        f     std         x           BodyGyro
59            fBodyGyrostdY        f     std         y           BodyGyro
60            fBodyGyrostdZ        f     std         z           BodyGyro
61          fBodyAccMagmean        f    mean                 BodyAccMag
62           fBodyAccMagstd        f     std                 BodyAccMag
63  fBodyBodyAccJerkMagmean        f    mean            BodyBodyAccJerkMag
64   fBodyBodyAccJerkMagstd        f     std            BodyBodyAccJerkMag
65     fBodyBodyGyroMagmean        f    mean               BodyBodyGyroMag
66      fBodyBodyGyroMagstd        f     std               BodyBodyGyroMag
67 fBodyBodyGyroJerkMagmean        f    mean           BodyBodyGyroJerkMag
68  fBodyBodyGyroJerkMagstd        f     std           BodyBodyGyroJerkMag
```

This is the code for creating that table. It just looks and changes particular values to things a little more easy to understand.

```
codebook <- data.frame(codes,timefreq= "", measure ="",direction = "", stringsAsFactors = FALSE)
codebook$timefreq[c(grep(codebook$codes, pattern = "^t"))] <- "t"
codebook$timefreq[c(grep(codebook$codes, pattern = "^f"))] <- "f"
codebook$direction[c(grep(codebook$codes, pattern = "X$"))] <- "x"
codebook$direction[c(grep(codebook$codes, pattern = "Y$"))] <- "y"
codebook$direction[c(grep(codebook$codes, pattern = "Z$"))] <- "z"
codebook$measure[c(grep(codebook$codes, pattern = "mean"))] <- "mean"
codebook$measure[c(grep(codebook$codes, pattern = "std"))] <- "std"
codebook$feature <- gsub("^[tf](.*)(mean|std)|","\\1",codebook$codes)
codebook$feature <- gsub( pattern = "(.*)[XYZ]$","\\1" ,codebook$feature)
write.table( codebook, row.name=FALSE ,file = "./codebook.txt")
```